

## US EPA TOXCAST DATA RELEASE OCTOBER 2014 - Database & R Package

This file describes the contents of the October 2014 ToxCast data release. The zip file contains the following summary-level files, not including this README file:

```
[1] "invitrodb_v1.mysql"
[2] "Pipeline_Overview.pdf"
[3] "tcpl_0.2.tar.gz"
[4] "tcpl_Package_Demo.html"
```

To utilize the database, the user must have write access to a MySQL server, allowing them to install the database contained in the "invitrodb\_v1.mysql" file. To avoid errors while calling large queries we recommend the MySQL server setting, 'max\_allowed\_packet', is set to at least 1 GB.

The "tcpl\_0.2.tar.gz" file is the beta version of the R package used to process all of the data in this release. The package can be installed using the following R command:

```
install.packages("tcpl_0.2.tar.gz", type = "source", repos = NULL)
```

To install the package on Windows the user will first need to properly install R tools, allowing R to compile the source package. The package relies on R (>= 3.1.0), and the following packages: data.table (>=1.9.4), DBI, RSQLite, MASS, numDeriv, RColorBrewer. To interact with the MySQL database, the user also needs the "RMySQL" package. The "RMySQL" package requires a local MySQL server, although it does not require the server to be running. More information about installing RMySQL on Windows can be found online.

The "Pipeline\_Overview.pdf" file contains slides illustrating the pipeline process at each level. It also includes appendix slides explaining the database field names for the level tables. The "tcpl\_Package\_Demo.html" walks through how to analyze an assay, and demonstrates some of the functionality of the R package.

For questions or concerns, please contact Monica Linnenbrink at: [linnenbrink.monica@epa.gov](mailto:linnenbrink.monica@epa.gov).